In *The Social Science Encyclopedia*, 3rd edition, ed. by Adam Kuper and Jessica Kuper, vol. 2, pp. 556-559. London: Routledge (2004)

LANGUAGE

Language is the most human of all human abilities. It may be the defining characteristic of *Homo sapiens.* Wherever humans exist, language exists. Although no one knows the precise number of languages in the world, something in the order of 6,000 to 6,600 is a good estimate, the exact number depending on one's definition of language versus dialect. (For the layman, the term 'dialect' often connotes a substandard deformation of a standard language or, sometimes, an unwritten language spoken by a small tribal group. As a technical linguistic term, 'dialect' simply refers to any socially or geographically recognizable variant of a language, e.g. Standard American English, Oxbridge English, Australian English, Indian English, Appalachian English, cockney and black vernacular English are all 'dialects of English'.) Considering that the world is populated by billions of people, the number of distinct languages is actually rather small. In addition, a large portion of the world's population speaks only a small handful of these thousands of languages: Chinese (Mandarin), English, Spanish, Portuguese, Russian, Arabic, Hindi and Indonesian/Malay count among the extra-large languages with hundreds of millions of speakers. English can rightfully be considered the world's most widespread language, especially when one takes into account second-language users, but it is not the language with the most speakers. Mandarin Chinese, with close to a billion speakers, can claim this first spot.

Although some 6,000+ distinct languages exist, most of them can be grouped into families of related languages in the same way that different plants and animals can be grouped into species, genera and families. There are a few isolates, such as Basque and Hadza (Tanzania), but these are rare. Better-known families are Indo-European, to which English belongs, Sino-Tibetan (East Asia), Athabaskan (North America) and Niger-Congo (Africa south of the Sahara). Lower-level families such as Bantu and Semitic are easily recognizable whereas historically much deeper and larger superfamilies such as Nostratic (Europe, Northern Asia, the Caucasus and the Middle East) and Amerind (almost all native languages of the New world from southern Canada to Patagonia) are more tenuous groupings that are subject to ongoing debate.

At first sight, the world's languages show remarkable and seemingly unlimited variation in appearance and structure. Apache, Chinese, Hebrew, !Kung, Lapp, Mohawk, Mongolian, Russian, Thai, Uzbek, Warlpiri and Zulu seem very different from one other in almost every respect. Yet, despite these surface differences, all human languages are governed by universal properties and constraints. This fundamental notion can be [557] traced back to thirteenth century observations by Roger Bacon; but in modern times it owes its recognition and scholarly impact to the ideas of Noam Chomsky, the dominant linguist and innovative thinker about language in the second half of the twentieth century.

A basic postulate held by all linguists, and taught in every introductory linguistics course, is that human languages are all equally complex. This is probably not literally true; certainly it has never been shown to be so. What does seem to be true is that there are no so-called primitive languages in the sense that all human languages are essentially equally capable of expression. If one can express a proposition in one language, the same thought can be expressed in another, although the form and manner of expression may differ considerably. The expressive potential of languages shows up in the remarkable elasticity of their lexicons. The vocabulary of every language is a living organism that is always in the process of being enriched to include new words for new things or concepts. Vocabulary expansion is accomplished through a variety of means such as borrowing words from another language, e.g. *pizza* (from Italian) or *hadj* (from Arabic), combining words to form compounds, e.g. *skyscraper* or *pickpocket*, blending words together, e.g. *smog* from *smoke* and *fog* or *motel* from *motor* and *hotel*, coining new words, a common practice of manufacturers of new products as exemplified by *aspirin*, *kleenex*, and *bandaid*, or using acronyms as single words, e.g., *radar* from *radio detecting and ranging, or AIDS* from *acquired immune deficiency syndrome*. Abbreviated forms of longer words or phrases may also become lexicalized, as exemplified by *ad* for *advertisement* or *bra* for *brassiere*, and proper names may be used as common terms, such as *champagne*, from the *Champagne* region in France (although the French object strongly to this generic usage or 'genericide') or *sandwich*, named from the fourth Earl of Sandwich, who, it is reported, ate his food between slices of bread so that he need not take time off from gambling to eat in normal fashion. These examples are all from English, but all languages can expand vocabularies in similar fashion, as is shown by compounds such as *cure-dent* (toothpick) in French, *Panzerkraftwagen* (armoured car) in

German, *farar-hula* (civilian, lit. white cap) in Hausa (Nigeria), or *heneba* (prince, lit. son of chief) in Akan (Ghana).

A universal characteristic of languages is that the form/sound of vocabulary items has no natural real-world connection to its referent or meaning. Most languages have a few items such as 'mama' or 'meow' that are onomatopoetic, i.e. there is supposedly some direct connection between the sound and the meaning; but the essence of human language is a system composed of totally arbitrary signs whose meaning is assigned only by convention and history. For example, the word meaning 'house' is *house* in English, *maison* in French, *casa* in Spanish, *dom* in Russian, *gida* in Hausa and *nyumba* in Swahili. In none of these languages, or any of the thousands of other languages of the world, is the pronunciation of the word any more indicative of the meaning 'house' than in any of the others.

All human languages utilize a finite set of discrete sounds like 's', 'm', 't', 'a' and 'i', which can be defined by a finite set of phonetic properties or features. In addition to consonants and vowels, accent (as in English) or tone (as in Yoruba) are common properties of the phonological inventory of languages. Some sounds, e.g. the clicks of !Kung and Xhosa and other southern African languages, are nonexistent elsewhere in the world, but most languages draw on the same finite universal set of phonological units.

The sound segments combine to form syllables that in term form meaningful units like *cat*, *play, common*, or *-ess* (as in *princess*) called *morphemes.* Some words consist of just one morpheme; others are complex morphological units in which simple morphemes combine to form words like *cats*, *replay*, *uncommonly* or *princess*. Each language has specific constraints on word formation. In English one can add *un-* as a prefix to negate the meaning of word, as in *unlikely* or *unfortunate*, but one cannot add it at the end as a suffix, i.e. *likelyun* and *fortunateun* are not words in English. Nor can one can one add it to all words even as a prefix, note the nonexistence of *uncat* and *unplay*.

Just as in word formation, there are constraints or grammatical rules that determine how words can be combined to form sentences. *The cat is on the mat* means something different from *The mat is on the cat*, and *cat the on is mat the* means nothing because the words are not combined according to the syntactic rules of English.

The syntactic rules in every language are similar in kind although they may differ in [558] specific constraints. Thus, in English, adjectives typically precede the nouns they

modify (as in *the red house*) whereas in French they usually follow (as in *la maison rouge*). But in all languages these rules of syntax include a principle of productivity and recursion which permits the generation of an infinite set of phrases and sentences. We know that this is so since any speaker of any language can produce and understand sentences never spoken or heard previously. This recursive aspect is also revealed by the fact that, in principle, there is no longest sentence in any language: one can keep adding additional words or phrases or conjoin sentences with words like *and* or *but* or relative clauses, such as *The cat is on the mat and the mat is on the floor*, or *The cat is on the mat that is on the floor*, or *The cat is on the mat and the mat is on the floor and the floor is made of wood that comes from the forest in the north of the country near the border that separates Maine from Canada.*

Speakers of a language know these rules. The system of knowledge that underlies the ability to speak and understand the infinite set of sentences constitutes the *mental grammar* of a language that is acquired by a child and is accessed and used in speaking and understanding. This underlying linguistic knowledge is not identical to the functional processes used in speaking and understanding. In actual linguistic performance, speakers access this mental grammar along with other non-linguistic systems (motor, perceptual, cognitive) in order to create and understand speech. This difference between the knowledge of language (the grammar) and linguistic performance accounts for why in principle language is boundless and sentences may be infinitely long, whereas in performance, where memory and other physical limitations come into play, each sentence is finite and the total number of sentences produced and understood in any one lifetime is finite.

The universality of language and of the grammars that underlie all languages suggests that the human brain is uniquely suited for the acquisition and use of language. This view is receiving increasing support from research on child language acquisition and from neurological studies of language disorders such as aphasia. No one now questions the position put forth by Paul Broca in 1861 that language is specifically related to the left hemisphere. Furthermore, there is converging evidence that focal damage to the left cerebral hemisphere does not lead to an across-the-board reduction in language ability, and that lesions in different locations in the left brain are quite selective and remarkably consistent in the manner in which they undermine language. This selectivity reflects the different parts of the grammar. Access to and processing of the phonology (sound system), the lexicon

(inventory of morphemes and words), the syntax (rules of sentence formation) and the semantics (rules for the interpretation of meanings) can all be selectively impaired. There is also strong evidence showing that the language faculty is independent of other mental and cognitive faculties. That is, language not only appears to be unique to the human species but also does not appear to be dependent on general intelligence. Severely retarded individuals can learn language, whereas persons with brain lesions may lose language abilities and still retain other cognitive abilities.

An aspect of language about which we know little but which is currently the subject of exciting and promising research is the origin of language. In the mid-nineteenth century, the question of language origin was viewed as so speculative and amateurish that the leading linguists of the day banned further debate on the subject. With recent developments in cognitive science and human paleontology and genetics, the subject has had a rebirth. The best way to provide a picture of where things now stand with regard to research on the origin of language is to focus on questions rather than answers (where consensus is still sorely lacking). Was language created once (monogenesis) — the majority view — or did different groups of early humans independently create language (polygenesis)? When did this happen? Estimates range wildly from as early as 150,000 –200,000 BP (coinciding with the appearance of *Homo sapiens sapiens*), to barely 50,000 years ago (a date that is probably much too recent). Did the language faculty appear suddenly almost full blown from rich mental capacities already in existence or did language first appear in very rudimentary form only to reach its current elaborate grammar through a long evolutionary process involving adaptive selection? Assuming an African origin for modern humans, what is the connection between the uniquely human capacity for language and the successful spread of modern man out of African some 60,000 years ago to the four corners of the world?

In the final analysis, the origin and spread of [559] language is not an esoteric historical question, but rather is a core question that will help us understand the striking feature of the 6,000 or so languages of the world that we know today, namely that they are incredibly, and often surprisingly, different from one another, and yet, when one probes deeply enough, they all seem to come from a common mould.

VICTORIA A. FROMKIN
University of California at Los Angeles
PAUL NEWMAN
Indiana University

Further Reading

Chomsky, N. (1975) *Reflections on Language*, New York.

_____. (1986) *Knowledge of Language: Its Nature, Origins, and Use*, New York.

Crystal, D. (1997) *The Cambridge Encyclopedia of Language*, 2nd. edn, Cambridge, UK.

Garry, J. and Rubino, C. (eds) (2001) *Facts about the World's Languages: An Encyclopedia of the World's Major Languages, Past and Present*. New York.

Pinker, S. , and Bloom, P. (1990) 'Natural language and natural selection (with commentaries)', *Behavioral and Brain Sciences* 13: 707-84.